DOCUMENT RESUME

ED 133 326 SP 010 727

TITLE Educational Accountability. Info-Item Educators Digest/No. 1020.

INSTITUTION National Education Association, Washington, D.C.; Ohio Education Association, Columbus. Instruction and Professional Development Div.

NOTE 65p.

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.

DESCRIPTORS Achievement Rating; Achievement Tests; *Educational Accountability; *Educational Trends; Evaluation Criteria; *Measurement Techniques; Performance Factors; Program Effectiveness; School Responsibility; *State Standards; *Teacher Evaluation; Testing Problems

IDENTIFIERS Ohio

ABSTRACT

Educational accountability is examined from the viewpoint of (1) history--the development of the concept from that of cost accounting to one of the educator's responsibility to parents and taxpayers; (2) new requirements of the Ohio state government for yearly reports on student skills in various subject areas according to sex, race, size of community, and socioeconomic status; and (3) new Ohio state requirements for progress reports for each school district and its constituent schools on plans, problems, personnel, and supplies. In addition, two papers by the National Education Association are reprinted: "Problems in Using Pupil Outcomes for Teacher Evaluation" and a comparison of random sampling and total population testing theories, "Why Should All Those Students Take All Those Tests?" The document concludes with an Ohio Education Association position paper on the uses and limitations of the accountability concept. (MB)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED133326

IPD
IPD
IPD
IPD
IPD
IPD
IPD
IPD
IPD
IPD

Info-Item
Educators Digest/#1020

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRO
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRE
SENT OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

Educational Accountability

SP010 227

distributed by:
Instruction and Professional Development Division
Ohio Education Association/National Education Association

OHIO EDUCATION ASSOCIATION

PRESIDENT: Sally Savage

EXECUTIVE SECRETARY: E. Wade Underwood

ASSISTANT EXECUTIVE SECRETARY: Robert E. Miner

INSTRUCTION AND PROFESSIONAL DEVELOPMENT DIVISION

DIRECTOR - ORGANIZATIONAL DEVELOPMENT AND RESOLUTIONS:

Edward F. Jirik
Eileen Batey - Secretary
Bobbi Ceisler - Conference Specialist

ASSOCIATE FOR INFORMATIONAL SERVICES AND AFFILIATED ORGANIZATIONS:

Byron H. Marlowe
Karen Kerr - Secretary
Judy Catalina - Information Specialist

ASSOCIATE FOR INVESTIGATORY SERVICES (PR&R and TEPS):

James F. Brehm
Kathy Newman - Secretary

ASSOCIATE FOR HUMAN RELATIONS AND AFFIRMATIVE ACTION:

Ramon H. MacFarlane
Sheila Valentine - Secretary

REPORT DIRECTOR: Edward F. Jirik

DISTRIBUTION: This booklet was printed by the Ohio Education Association as an informational item for local association leaders.

Educational Accountability

TABLE OF CONTENTS

SECTION	PAGE
Introduction.	1
State Testing Programs.	2
Ohio "NEW" Annual Educational Assessment Standards.	3 - 7
Ohio "NEW" Annual Progress Report Guidelines.	8 - 22
Problems In Using Pupil Outcomes for Teacher Evaluation. . . .	23 - 41
Why Should All Those Students Take All Those Tests?.	42 - 50
The OEA Accountability Position Paper.	51 - 61

Educational Accountability

Introduction

The term *accountability* was first used in regard to education in 1969 when Leon Lessinger, as U. S. Department of Health, Education, and Welfare Associate Commissioner of Education, came up with an idea that seemed as reasonable as it was novel - - that grant seekers should specify precisely their intended educational outcomes and costs of their projects.

Those receiving grants were to be audited to see whether they had indeed achieved these outcomes within the specified costs.

This rather limited concept expanded to become much broader in meaning, as is evidenced in this definition by Lessinger, Parnell, and Kaufman:

> Accountability in education means just what its dictionary definition says it means: responsibility. If you are held accountable for something, you are responsible for it, answerable to someone about it. In education, accountability means that educators of all kinds should be answerable to parents for how effectively their children are being taught and answerable to taxpayers for how usefully their money is being spent.[1]

To have a complete view of the meaning that surrounds accountability we must examine yet another way of looking at the concept. Many seem to see accountability as synonymous with the methods employed to achieve it . For example, in the past when many educators spoke of accountability they meant performance contracting. Other writers and educators may actually be referring to things like merit salary programs, Jencks' voucher plan, or systems management techniques like PPBES. It must be pointed out that these systems are merely methods; they do not define accountability but are, as Lessinger and his associates pointed out in a 1973 volume, merely *tools* for the achievement of accountability.

Those definitions reflect the differences in people's ideas about *what effective education is;* as long as educators continue to argue this issue they will continue to disagree about the definition of educational accountability. It will be up to teachers and other educators to formulate the definition as we learn more and more about our educational responsibilities to children and how to achieve them.

---

[1]Lessinger, Leon M.; Parnell, Dale; and Kaufman, Roger. "Learning." Volume I of Accountability: Policies and Procedures. (A Series of 4 Volumes.) [New London, Connecticut]: Croft Educational Services, 1971.

State Testing Programs

Some 30 states have enacted a form of accountability legislation. Of the 30 states that are now required by law to implement accountability programs, 18 have enacted state testing programs. Still others have enacted programs utilizing testing. Standardized testing is specified by law in at least nine of these programs.

There is very little information available about the details of state accountability programs utilizing standardized testing. Indeed, it is difficult to ascertain whether these programs are being implemented at all. It seems likely that although everybody is talking about accountability, very few people are doing anything about it, or at any rate, many who are doing something about it aren't talking.

At least 13 states now use criterion-referenced tests in their statewide assessment programs, and there are indications that more may soon follow suit. Accountability programs using this type of testing are somewhat better reported than programs using standardized tests. Three of the most widely publicized programs are in Florida, California and Michigan.

The Florida program, utilizing both criterion-referenced and norm-referenced tests, is based on Florida's 1971 Educational Accountability Act. The criterion-referenced component of the testing has thus far been devised by Florida reading specialists and teachers who chose performance objectives from a catalog provided by the Center for the Study of Evaluation at the University of California at Los Angeles. The program, projected through 1978, includes plans to measure student performance in such diverse areas as mental health and aesthetic appreciation as well as communication and learning skills.

The California program is based on the 1972 Stull Act, which requires each teacher to develop pupil performance objectives and criterion-referenced tests as a basis for evaluation of his or her work. In 1972-73 the San Diego Unified School District responded to the act with a plan prepared by teachers and principals for teacher evaluation based on student performance on certain learning objectives. Although a few other similar kinds of programs have been instituted in California, it is unclear what kinds of programs most schools in the state are instituting, or indeed if they *are* instituting serious programs. A paper from the Institute for the Development of Educational Activities notes, regarding California, that "teachers and administrators consider that state's accountability program 'a paper tiger'."

The Michigan program, begun in 1970, is one of the pioneer state accountability programs. It originally utilized norm-referenced tests but after two years replaced them with criterion-referenced tests developed by the state board of education, teachers, and administrators. At present the program measures performance only in reading and math, but plans are being made for testing in other areas. In the future, the state plans to avoid spending the millions of dollars necessary to test all students by testing only a representative sample of students on most objectives. A 1974 National Education Association-sponsored evaluation of this program severely criticized it for using performance objectives that purportedly were not field-tested or validated and that penalized minority students. The NEA committee recommended the use of local rather than statewide objectives.

Ohio 'New' Annual
Educational Assessment Standards

ANNUAL EDUCATIONAL ASSESSMENT STANDARDS

INTRODUCTION

In accordance with Section 3301.131 of the Ohio Revised Code, the Department of Education shall administer an Annual Educational Assessment Program. The program shall require studies to be conducted to provide data from samples of Ohio students in a variety of subject and skill areas, which shall include reading and English composition at several grade levels. Data collected shall be used to identify schooling needs, while preserving the anonymity of students, teachers, schools, and school districts. The assessment program shall assure data from all students tested, including students from racial and cultural minority backgrounds.

EDb 103.01 SCHEDULE OF ASSESSMENT SUBJECTS AND GRADE LEVELS

I. The subjects and grade levels shall be assessed for the first two years according to the following schedule:

   A. During the 1976-1977 school year, reading shall be assessed in grades 4, 8, and 12; English composition in grade 8; mathematics in grade 8.
   B. During the 1977-1978 school year, English composition shall be assessed in grade 12 and mathematics in grades 4 and 12.

II. The subjects and grade levels to be assessed subsequent to the 1977-1978 school year shall be determined by the Ohio Assessment and Annual Report Advisory Committee established in Section 3301.131 of the Ohio Revised Code.

EDb 103.02 ASSESSMENT INSTRUMENTS, OBJECTIVES, ITEMS AND DESIRED STUDENT PERFORMANCE LEVELS

I. Assessment instruments shall be based on performance objectives appropriate for each grade level and subject. The length of instruments shall be appropriate for each subject and grade level.

II. Performance objectives shall be selected with the advice of members of the education profession. Objectives selected shall be appropriate for the grade level to be assessed, represent knowledge which is an important part of a student's learning, and represent what is or should be taught in Ohio schools. Objectives selected for each subject and grade level shall be of a basic nature, reflecting skills and knowledge important for all children in their daily lives.

III. Each assessment item shall be appropriate for one or more of the identified objectives, and may be newly created or obtained from other sources. Each item shall be reviewed by representatives of the education profession, including testing experts, child development specialists, teachers and subject matter specialists to insure that it is a direct measure of the objective to be assessed, appropriate for the grade level, clearly written, free of sex, racial or cultural bias, and presents situations or problems similar to those that students are likely to encounter in everyday life. Multiple choice or other types of objective items shall be examined to insure that only one correct answer is presented among choices.

IV. Representatives of the education profession shall review items in the instruments and determine the level of performance desired for students being assessed. Schooling needs shall be determined by comparing such desired levels with actual levels of student performance, as indicated by results on assessment instruments.

EDb 103.03 SAMPLING DESIGN

I. The assessment program shall use a random sampling procedure for each subject area and grade level, and shall generate a sample of students of sufficient size to yield results at reasonable and technically acceptable confidence levels. All students at each grade level to be assessed in Ohio public and nonpublic schools shall constitute a universe for the sample, excluding students who may be classified as hearing impaired, visually impaired, or students with severe and/or multiple impairments.

II. The sample design shall insure that data are obtained for each of the following sampling categories:

A. Sex
B. Race and ethnic composition, including:

1. Black/Negro
2. White/Caucasian
3. American Indian or Alaska Native
4. Asian or Pacific Islands
5. Hispanic

C. The sample design shall be so constructed as to allow the following report categories:

1. Sex (male, female)
2. Race (Black, White)
3. Size of community (four population size categories)
4. Socioeconomic status (low-medium-high) of students
5. Race by socioeconomic status

EDb 103.04 PROGRAM ADMINISTRATION

I. Annual administration of the assessment instruments shall occur no earlier than March 1 nor later than April 30 of each year.

II. School buildings selected to participate in the assessment program shall be provided instruction adequate to administer the assessment instruments and to insure that students are selected on a random basis in accordance with proper sampling procedures. Such instructions shall emphasize the statewide nature of the assessment program. The administration of the program shall insure the procedures for random selection of students are followed.

III. The assessment program shall be conducted in such a way as to assure the anonymity of students, teachers, buildings, and districts.

EDb 103.05 FORMAT AND DISTRIBUTION OF ASSESSMENT RESULTS

I. The results of the state assessment program shall be summarized by the Ohio Assessment and Annual Report Advisory Committee and such summary shall preface or cover any assessment data released. Annual assessment report summaries shall be distributed to all public and nonpublic school districts, school buildings, members of the General Assembly, the State Board of Education, and the Ohio Assessment Advisory Committee and those individuals and groups assisting with each annual assessment program. Varying levels of detail may be reported to different groups according to differing needs.

II. Assessment reports shall contain data reported by objectives and by related sampling variables, and shall be prepared in such a manner as to be useful to the education profession and citizens.

III. Subsequent to the administration of assessment instruments, a substantial number of objectives and items, along with definitions of the reporting categories, shall be made available to school districts on request for their use.

Ohio 'New' Annual Progress Report Guidelines

ANNUAL PROGRESS REPORT GUIDELINES

INTRODUCTION

In accordance with Section 3313.94 Ohio Revised Code, each school district shall prepare and distribute annually a written progress report for each of the schools under its jurisdiction and for the total district. The first annual progress report shall be prepared and distributed no later than November 1, 1976, or such later date as the Ohio General Assembly may authorize, with subsequent reports prepared and distributed on or before November 1 of each successive year. The reporting period shall be the preceding school year (July 1-June 30).

The guidelines, delineated below, are minimum; they are not intended to be restrictive, rather they are guiding, suggestive and directive. Schools are encouraged to go beyond the requirements.

EDb 411.01 DEFINITIONS

Data to be reported in the Annual Report shall utilize the following definitions:

I. For standards EDb 411.02 through 411.05:

A. Achievements are measurable student accomplishments. Achievements may include, but are not limited to, the reporting of curricular goals attained, awards/recognition, test scores or post high school graduation data about students, etc. If test scores are utilized, appropriate testing procedures which recognize the unique differences of special education students should be employed and provisions should be made to keep the data separate; pre and post test information should be used; student aptitude should be compared to achievement and results should be presented on the basis of student ability.

B. Problems are either specific shortcomings in programs or special factors bearing adversely upon the operation of schools. Problems may include, but are not limited to, reasons for failing to achieve goals on schedule, inadequate financing, low attendance, vandalism, violence or low level of community support.

Problems are things that stand in the way of attaining the basic objectives of each district and school.

C. Plans are steps for attaining stated goals. They may include, but are not limited to, provisions for research, special assignment of personnel, use of outside resources and special costs.

D. Improvements are the actual results achieved in attaining specific goals. Improvements may include, but are not limited to, new programs, higher rates of attendance, fewer dropouts, reduced vandalism, new or remodeled facilities, expanded media centers, laboratories, shops, or classrooms, etc.

II. For standards EDb 411.02-II-A-2 and EDb 411.05-II-A-2:

A. Teaching Personnel includes librarians and media specialists, aides, speech therapists, and substitutes.

B. Pupil Services include guidance counselors, physicians, nurses, psychologists, and visiting teachers/social workers.

C. Central Office Personnel includes superintendents, assistant superintendents, directors, supervisors, and coordinators or similar classifications as appear in the Ohio Education Directory.

D. General Services and Supplies include fixed charges for rent, taxes, assessments, dues to educational associations, auditor and election costs, expenses of board members, legal and other special services, recreation, and other miscellaneous expenditures and fees.

E. Teaching Services and Supplies include textbooks, workbooks, teaching supplies, library supplies, library and audio-visual materials, playground and community centers, travel of instructional staff, and salaries for part-time teachers.

F. Maintenance includes salaries, supplies, and contract services.

G. Transportation includes salaries, supplies, and contract services.

H. Utilities include electricity, water, telephone, and fuel (gas, oil, coal).

III. For standards EDb 411.02 through 411.05:

A. Classroom teacher is a certificated staff member who has direct contact with students for the purpose of instruction. A teacher may serve in a regular classroom or as a special resource teacher. At the elementary level the information should agree with the data in Section V of the Elementary School Principal's Annual Report. At the high school level the data reported should be consistent with the information in Section IV of the Secondary Principal's Annual Report.

1. A regular classroom teacher is a teacher who meets with a group of students on a regularly scheduled basis. A regular classroom teacher generally has a homeroom and is responsible for recording and reporting information to parents of those homeroom students.

2. A special education teacher is a regular classroom teacher who teaches a class which meets the Special Education Standards of the Department of Education.

3. A vocational education teacher is a regular classroom teacher who teaches a class which meets the Vocational Education Standards of the Department of Education.

4. A special resource teacher is a teacher who meets with students but does not fit one of the three categories above. At the elementary level examples of a special resource teacher could be remedial reading teacher, librarian, foreign language teacher or a teacher teaching special remedial or enrichment courses. A special resource teacher would vary from building to

building depending upon the course of study and the school organization. At the secondary level, a special resource teacher would be one who teaches classes for which no credit is given.

B. Administrative, Supervisory, and Pupil Personnel staff include certificated staff not meeting with students for the basic purpose of instruction. Central office personnel as defined in section II, school psychologists, principals, guidance counselors, etc., are included in this category.

C. Student mobility is the ratio, in percentage, of the total number of transactions on the enrollment records of an individual building to the average daily membership of that building. Transactions include the beginning enrollment, new entries to the building, transfers or withdrawals from the building and reentries. Expulsions or suspensions would not be included. Transactions which occur between the opening of school and the first full week of October should not be included.

Edb 411.02 CITY, EXEMPTED VILLAGE, AND LOCAL DISTRICT REPORTS

I. Each school district shall devise a procedure for releasing to the public on a yearly basis data indicating the achievements, problems, plans and improvements made by the district in meeting the educational goals established by either: (a) the State Board of Education, (b) the district board of education, or (c) the individual buildings.

This narrative report may deal with the areas of attendance, budget, curriculum, extra curricular activities, facilities, staffing, staff development, special services, transportation, etc.

II. In addition to the preceding information, the Annual Progress Reports for each city, exempted village and local school district shall include the following statistical information:

A. Financial Data, including:

1. Average Per Pupil Costs. Percentage of total expenditures from local tax revenue and percentage of expenditures from other revenue sources.

2. For the first report, total dollars and percentages of general fund operating expenditures in either: (a) the six major reporting categories as designated in the Department of Education's publication entitled "Cost Per Pupil," or (b) the following categories:

1. School Building Personnel
   - Teaching
   - Principals and Assistants
   - Pupil Service
   - Custodial

2. Central Office Personnel
   - Instruction
   - General Administration
   - Finance and Business

3. General Services and Supplies

4. Teaching Services and Supplies

5. Maintenance

6. Transportation

7. Utilities

If the latter method is used, the definitions in EDB 411.01-II should be utilized along with any special instruction provided by the Department of Education. In subsequent years, the method outlined in (b) shall be utilized showing two year trends.

3. Information on interest, debt retirement, motor vehicles and capital outlay expenditures as defined in the Department of Education "Cost Per Pupil" Report.

B. Staff Data, including:

1. Number of full-time equivalent regular, special education and vocational education teachers per 100 students.

2. Number of full-time equivalent special resource teachers per 100 students.

3. Number of full-time equivalent administrative, supervisory and pupil personnel staff per 100 students.

4. Number of full-time equivalent paid instructional aides per 100 students.

5. Average classroom teacher experience and training based on data reported in Form SF-1, "Report of Certificated Employees."

6. Average classroom teacher salary based on data reported in Form SF-1, "Report of Certificated Employees."

7. Percentage breakdown of teachers and administrative, supervisory and pupil personnel staff by race based on data reported in Form SF-1, "Report of Certificated Employees."

8. Percentage breakdown of teachers and administrative, supervisory and pupil personnel staff by sex based on data included in school district records.

C. Pupil Data, including:

1. Enrollment figures including race and sex data.

2. Average daily attendance based on data reported in Form SF-2, "Report of Certified Average Daily Membership."

EDb 411.03 BUILDING REPORTS

I. Each building, in accordance with the procedure established by the district board of education, shall report to the public on a yearly basis, data indicating the achievements, problems, plans and improvements made by the building in meeting the educational goals established by either: (a) the State Board of Education, (b) the district board of education or (c) the individual building.

This narrative report may deal with the areas of attendance, budget, curriculum, extra curricular activities, facilities, staffing, staff development, special services, transportation, etc.

II. In addition to the preceding information, the Annual Progress Reports for each building shall include the following statistical information:

A. Staff Data, including:

1. Full-time equivalent regular teacher, special education teacher and vocational teacher/pupil ratio.

2. Number of full-time equivalent special resource teaching staff.

3. Number of full-time equivalent administrative, supervisory and pupil personnel staff.

4. Number of full-time equivalent instructional aides.

5. Number of full-time equivalent volunteers.

6. Average classroom teacher experience and training based on data reported in Form SF-1, "Report of Certificated Employees."

7. Percentage breakdown of teachers and administrative, supervisory and pupil personnel staff by race based on data reported in Form SF-1, "Report of Certificated Employees."

8. Percentage breakdown of teachers and administrative, supervisory and pupil personnel staff by sex based on data included in school records.

B. Pupil Data, including:

1. Total enrollment, including (a) sex of students, as reported in the Elementary Principal's Report Form-22 or the Secondary Principal's Report Form-23, whichever is applicable, and (b) race of students, as reported in forms collected through the Ohio Educational Data System.

2. Average daily attendance for the same period reported in Form SF-2 for the total district.

3. Student mobility.

4. Total expenditures for the past three years for library-media center materials including library books, periodicals, newspapers and nonprint materials. When a central library serves the particular school only, information describing the total times students used the library-media center during the past year should be presented. This may include class use of the facilities; individual use of the facilities; the number of times materials are circulated for student use, etc.

5. Economic status. One or more of the following indices may be reported: rate of unemployment, percentages receiving public assistance, percentages receiving free or reduced price lunches.

III. Annual Progress Reports shall include a statement desscribing the standardized test data which are available on individual students in the building and describe the procedure by which parents may review those data for their own children.

EDb 411.04 COUNTY BOARD REPORTS

I. Each county board of education shall devise a procedure for releasing to the public on a yearly basis data indicating the achievements, problems, plans and improvements made by the district in meeting the educational goals established by either: (a) the State Board of Education, or (b) the county board of education.

This narrative report may deal with the areas of attendance, budget, curriculum, staffing, staff development, special services, transportation, etc.

II. In addition to the preceding information, the Annual Progress Reports for each county school district shall include the following statistical information:

A. Financial Data, including:

1. Percentage breakdown of total expenditures showing contributions from local districts as well as state, federal and other resource agencies. Local district revenue should be identified by the district from which it is derived.

B. Staff Data, including:

1. Number of full-time equivalent administrative, supervisory and pupil personnel staff.

2. Average staff experience and training based on data reported in Form SF-1, "County Report of Certificated Employees."

3. Average staff salary based on data reported in Form SF-1, "County Report of Certificated Employees."

4. Percentage breakdown of staff by race.

5. Percentage breakdown of staff by sex based on data included in school district records.

C. If direct instructional services are provided to students, pupil data, including:

1. A description of the services provided.

2. Enrollment figures including race and sex data.

3. Cost per pupil of direct instructional services.

EDb 411.05 JOINT VOCATIONAL SCHOOL DISTRICT REPORTS

I. Each joint vocational school board of education shall establish a procedure to be utilized to report to the public on a yearly basis data indicating the achievements, problems, plans and improvements made by the district to meet the educational goals established by either: (a) the State Board of Education or (b) the district board of education.

This narrative report may deal with the areas of attendance, budget, curriculum, extra curricular activities, facilities, staffing, staff development, special services, transportation, etc.

II. In addition to the preceding information, the Annual Progress Reports for each joint vocational school shall include the following statistical information:

A. Financial Data, including:

1. Average Per Pupil Costs. Percentage of total expenditures from local tax revenue and percentage of expenditures from other revenue sources.

2. For the first report, percentages of general fund operating expenditures in either: (a) the six major reporting categories as designated in the Department of Education's publication entitled "Cost Per Pupil," or (b) the following categories and subcategories:

1. School Building Personnel
   - Teaching
   - Principals and Assistants
   - Pupil Service
   - Custodial

2. Central Office Personnel
   - Instruction
   - General Administration
   - Finance and Business

3. General Services and Supplies

4. Teaching Services and Supplies

5. Maintenance

6. Transportation

7. Utilities

If the latter method is used, the definitions in EDb 411.01-II should be utilized along with any special instruction provided by the Department of Education. In subsequent years, the method outlined in (b) shall be utilized showing two year trends.

3. Information on interest, debt retirement, motor vehicles and capital outlay expenditures as defined in the Department of Education "Cost Per Pupil" Report.

B. Staff Data, including:

1. Full-time equivalent teacher/pupil ratio.

2. Number of full-time equivalent administrative, supervisory and pupil personnel staff.

3. Average teacher experience and training based on data reported in Form SF-1, "Report of Certificated Employees."

4. Percentage breakdown of teachers and administrative, supervisory and pupil personnel staff by race based on data reported in Form SF-1, "Report of Certificated Employees."

5. Percentage breakdown of teachers and administrative, supervisory and pupil personnel staff by sex based on data included in school records.

C. Pupil Data, including:

1. Total closing enrollment, including sex data, as reported in the Vocational Education Closing Report Form VE-22.

2. Total closing enrollment, including race data, as reported in the Vocational Education Closing Report Form VE-22.

3. Average daily attendance for the same period reported in Form SF-2 for the total district.

4. Number of students reported on Form VE-23 who are available to work and are working full-time, part-time, not working or whose status is unknown.

5. Number of academic courses offered at the joint vocational school; number of students receiving their full academic preparation at the joint vocational school.

6. Economic status. One or more of the following indices may be reported: rate of unemployment, percentages receiving public assistance, percentages receiving free or reduced price lunches.

EDb 411.06 REPORT FORMAT AND DISTRIBUTION

I. Annual Progress Reports for city, county, exempted village, joint vocational schools, and local districts shall be made available in printed form and in such language as to be useful to citizens. The report shall be issued to the local community of each district. Districts shall be permitted maximum flexibility in the manner of distribution. Reports may be published in the local news media, be on file in the school library, be sent home with each student, etc.

II. Annual building reports shall be made available in printed form and in such language as to be useful to citizens. The report shall be issued to the local community of each building. Buildings shall be permitted maximum flexibility in the manner of distribution. Reports may be published in the local news media, be on file in the school library, be sent home with each student, etc.

Problems in Using Pupil Outcomes for Teacher Evaluation

Robert S. Soar
Foundations of Education
Institute for Development of Human Resources
University of Florida

Ruth M. Soar
Florida Educational Research and Development Council

published by

NATIONAL EDUCATION ASSOCIATION
1201 16th Street, N.W., Washington, D.C. 20036

April 1975

FOREWORD

Increasingly state association leaders and staff, local teacher groups, and UniServ directors find themselves having to deal with the problems of inappropriate teacher evaluation. As most teacher leaders are aware, one aspect of such problems is reflected in attempts to evaluate teachers on the basis of student achievement.

The attached paper provides a combination of responses to these problems--research findings, technical problems in using test scores, and other considerations. It has been prepared by two nationally eminent researchers in the field. Citing the Soars can serve to increase the credibility of arguments against the use of student achievement for evaluating teachers. Their examples should be particularly useful in dialogues with school district research directors, testing and evaluation coordinators, and other administrators who are committed to using a year's growth in a year as a measure of teacher competence.

In some other NEA material we have called to attention the major reasons why teachers must not be evaluated on the basis of student achievement. Those reasons complement and are supported by the Soar paper and seem worth repeating here in that context:

1. The tests themselves are inadequate for such purposes. Banesh Hoffman has put it well:

> There is no generally satisfactory method of evaluating human abilities and capabilities.... Rough superficial evaluations are of course possible.... But the detection and evaluation of other than superficial ability is inevitably an art demanding insight, taste and knowledge. Current attempts to reduce it to a science and then mechanize it are not only dangerous but in a profound sense unscientific.[1]

[1]Hoffman, Banesh. "Psychometric Scientism." Phi Delta Kappan 48: 381; April 1971.

2. The nature of student populations is so varied that outcomes are often more influenced by those variables than by what teachers do. Gene Glass, noted researcher, reminds us:

> Nothing short of random assignment of pupils to teachers as an iron-clad administrative necessity will ensure that the teachers were in a fair race to produce pupil gains.[2]

3. Many of the conditions which measurably affect learning outcomes are conditions over which teachers have little or no control and they vary widely among schools. Among them are: the number of students teachers must work with each day; time available to teach; planning time; up-to-dateness of curriculum; appropriateness of materials and media: students' physical and emotional readiness for learning; opportunity for teacher in-service education; and most important, decision-making power on curriculum matters.

Each of the reasons cited is considered in one form or another in the Soar paper. And even though some of the technical explanations may go beyond the needs of teacher leaders in responding to the issues, they serve as backups to commonly held teacher association positions.

--Bernard H. McKenna
Professional Associate
NEA Instruction and Professional Development

---

[2]Glass, Gene V. "Statistical and Measurement Problems in Implementing the Stull Act." Mandated Evaluation of Educators: A Conference on California's Stull Act. (Edited by N. L. Gage.) Washington, D. C.: Education Resources Division, Capital Publications, 1973. p. 54.

PROBLEMS IN USING PUPIL OUTCOMES FOR TEACHER EVALUATION

During the past few years there has been mounting pressure for measuring the outcomes of education, with movement toward holding the teacher, the school, and the school system "accountable" for producing the student learning expected by society. Decreasing enrollments, tighter budgets, and a general trend toward cost effectiveness have added to the pressure.

Measuring pupil achievement has increasingly been proposed as a way of assessing the effectiveness of teaching, and in fact has been mandated by a number of states. This approach is superficially reasonable and attractive, but it is fraught with problems which have not been generally recognized.

H. L. Mencken once commented, "There's always a well-known solution to every human problem--neat, plausible and wrong." The use of pupil achievement as a way of evaluating the teacher, the school, or the school system embodies this misleading simplicity. The solution seems so straightforward: If the job of the teacher is to promote learning in pupils, then it seems reasonable to evaluate the teacher in terms of the amount of learning he produces in his pupils.

The parallel with the industrial setting is clear: If the job of the worker is to assemble relays, then it seems reasonable to count the number

of relays the worker assembles and pay him or her accordingly. But in applying this procedure to teaching, a number of problems emerge which have not been widely recognized. The relay assembler receives parts which are identical (at least within very close limits) on which he or she performs a prescribed set of operations, also identical. Then the completed units leave the assembler, again almost identical from one to another.

But none of this is true for the teachers. Pupils appear in the classroom differing in ability, level of achievement, home background, interest, motivation, age--differing in numerous ways. The teacher must recognize these differences as he or she strives to help individual pupils grow toward their own potential. Consequently, the teaching process will differ from pupil to pupil. If the teacher has been successful, each pupil will have improved educationally when he or she leaves the classroom but each will probably be no more like the others than when the year began.

A major dimension, then, of the problem of evaluating teachers in terms of pupil outcomes is the recognition that what goes on in the classroom is not the only, or the most powerful, influence on where a pupil stands in achievement at the end of the year.

Influences Other Than the Classroom

Research has shown that the differences pupils bring with them when they enter the classroom have significant influence on achievement.

Entry level ability (pretest or fall score) and socioeconomic status are major determiners of what a pupil's standing will be at the end of the school year. These influences probably are more widely accepted than any other, but they are highly interrelated so that one overlaps the other. In practice they cannot be effectively separated.

The fact that IQ and achievement scores in the fall are highly related to spring achievement scores is widely accepted but seldom documented. In a study of 81 fifth-grade classes, Soar and Soar (1973) found correlations between class averages (means) for fall IQ and spring achievement ranging from +.85 to +.90, and correlations between fall achievement and spring achievement ranging from .75 to .85. So the evidence is that as much as 80 percent of the variation in class averages for pupil achievement at the end of the year can be accounted for by pupil characteristics which existed at the beginning of the year, characteristics over which the teacher has no control.

The most extensive data on the influence of socioeconomic status on pupil achievement were presented in the Coleman Report, and more recently and more widely reanalyzed by Mosteller and Moynihan (1972) and Mayeske, et al. (1972). The studies show that as much as 80 percent of the variation in pupil achievement across schools (equal to a correlation of about +.90) can be accounted for by these factors.

Beyond these major influences there are others which help account for differences in pupil achievement and which should be considered. Although the research on family attitudes and support for learning in the home is not as extensive as that for pupil ability (pretest) and social status, it is consistent in indicating relationships between the educational values held by parents and their children's achievement in school. Garber and Ware (1972) found a relation of +.47 between achievement and a combined measure of support for learning in the home for a group of Black and Spanish-American children. All students in the sample met federal poverty guidelines, so that socioeconomic status as usually measured was, in effect, held constant. The same authors cite similar findings from other studies.

Peer group attitude, although again the research is not extensive, has been identified as another important factor which can either support or hinder a pupil's achievement (Anderson, 1970).

Since there is compelling evidence that a number of influences over which the teacher has no control have powerful effects on pupil achievement, it cannot be expected that a teacher will have consistent results with successive groups of pupils. That is, the teacher will not be equally effective in producing growth with all groups because groups differ so widely. Studies by Rosenshine (1970) and Brophy (1972), for example, show that on the average only about 10 to 15 percent of the variation in achievement from group to group reflects the stable influence of the teacher, as shown by a median correlation in the low 30's.

As Medley (1974) has pointed out, and as commonly accepted methods of estimating reliability show (Chronbach, 1960, p. 131), data from about twenty classes would be required for making reliable decisions about individual teachers. Given this requirement necessitating collection of such large amounts of data, using the measurement of pupil achievement as a way to evaluate teachers is impractical as well as invalid.

What these findings seem to indicate is that the education of the pupil is dependent on many conditions in the society, not on the school alone. When the time the pupil spends in the classroom is compared with the time he spends under other influences, and when the degree of influence or control the teacher can exercise is compared with the power of other influences, the limited effect of the teacher is not surprising.

Because influences other than the teacher make a major difference in how much the child learns is not to say that the role of the teacher is

unimportant. The teacher is the only formal, institutionalized input the society has to the education of the child and the transmission of an established curriculum. And much of what the teacher does that contributes constructively to the child's future abilities, successes, and satisfactions may not be measured by currently common achievement instruments. It does say, however, that the teacher's influence is limited and that the teacher is most effective when he or she has the support of other elements in the society.

This whole constellation of other influences is usually not given consideration when measures of pupil achievement are proposed as the basis for evaluating teachers. It is reasonable that these influences are strong, since they accumulate over the life of the pupil. It is obvious, then, that pupil standing at the end of any school year is a completely inadequate and even misleading measure of the effectiveness of the teacher or the school. Yet the results of such achievement standings are frequently published by school or by school system.

"Standing" versus "Change" as Measures of Outcome

"Achievement," which is the most frequently used measure of student learning outcomes, usually refers to the amount of knowledge a pupil possesses at a given point--his or her "standing." The influences cited above show a strong relation to achievement as used in this sense.

An alternative to measuring achievement standing is to measure "change" in achievement from the beginning to the end of the year. When this is done, the influences cited are still likely to have an effect, although to a lesser degree, since change reflects their influence for a shorter period of time.

Although this alternative is appealing as another way of evaluating teaching, it raises still other problems. In a classic volume on the problems

of measuring change, Bereiter (1963) commented:

> Although it is commonplace for research to be stymied by some difficulty in experimental methodology, there are really not many instances in the behavioral sciences of promising questions going unresearched because of deficiencies in statistical methodology. Questions dealing with psychological change may well constitute the most important exceptions. It is only in relation to such questions that the writer has ever heard colleagues admit to having abandoned major research objectives solely because the statistical problem seemed to be insurmountable. (p.3)

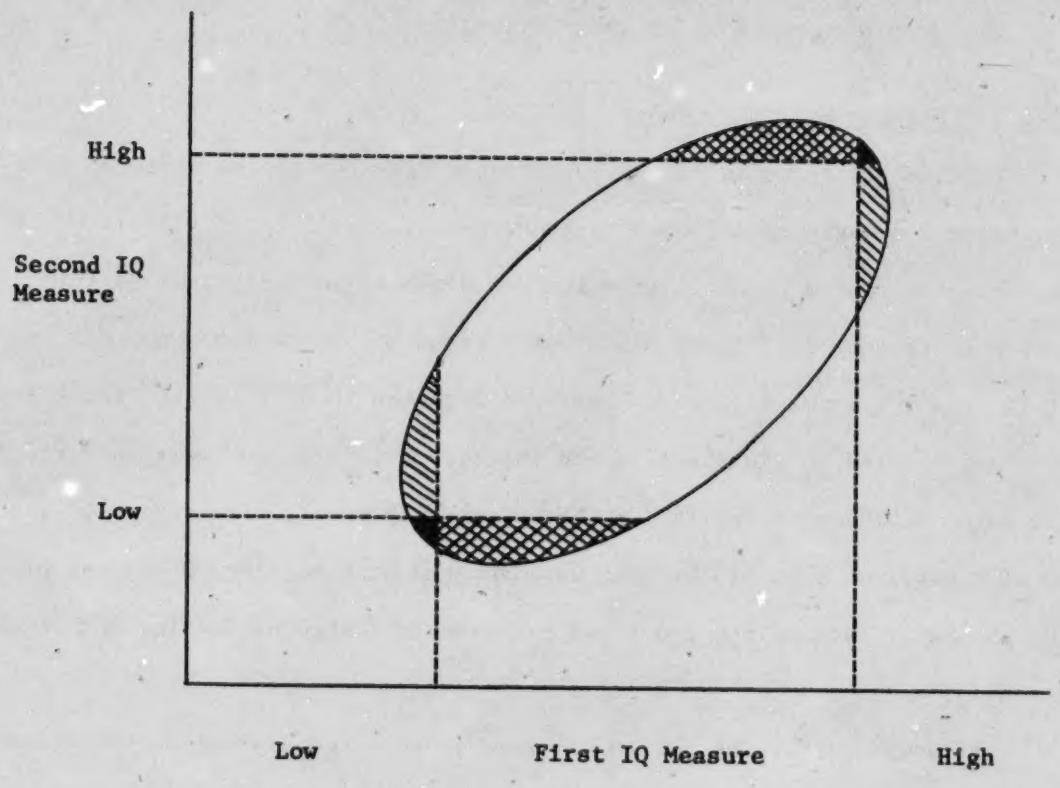
Difficulties in Measuring Change

If the fall score is simply substracted from the spring score so as to obtain a measure of net change, a new set of subtle but difficult problems is created. An illustration may serve to identify some of them. Figure 1 presents fictitious data from a group of pupils for whom measures of IQ from two forms of a test have been obtained 10 days apart. The initial IQ's are plotted on the baseline and the second IQ's on the vertical axis. Any point in the area outlined by the elipse represents simultaneously the IQ of a pupil on each of the testings, and the high and low 10 percent of the pupils at each of the two times has been indicated by shading and cross-hatching.

It is clear that the pupils who were in an extreme group on the first test were not, for the most part, in an extreme group on the second test. The blackened areas represent the small number of pupils who were extreme on both occasions.

At the upper right, the area is small because the pupils who make the highest scores at any testing are likely to do so on two bases: (1) they are bright (have high verbal skills), and (2) they are "lucky" (that is, they

Figure 1.--An Illustration of Regression Effect



happen to make good guesses on a few items for which they aren't sure of the answer, or the items on this test just happen to be ones for which they know the answers). But they are not likely to be lucky consistently when another form of the test is given, and so on another testing their scores are likely to be lower. Opposite influences will affect pupils at the lower left end of the elipse.

To put it another way, if the cutting point for the top 10 percent is an IQ of 120, there will be a number of pupils with true IQ's close to 120 who will sometimes be above that score on a series of tests and sometimes below it, depending on chance factors. So some fraction of pupils above 120 on the first test will fall below it on the second. Similarly, some fraction of the pupils scoring below 80 on a first test will be above it on a second.

In both cases, extreme pupils have "regressed," or moved, toward the mean. This regression effect can be expected whenever prediction is less than perfect, and the extent of the movement will depend on the inaccuracy of the prediction (Lord, 1963). With most psychological or educational predictions, the regression involved is considerable and may make up a significant proportion of the total range of scores.

The point to be stressed from this example has important consequences: Since pupils who were in the bottom 10 percent the first time were not, for the most part, in that group the second time, they must have moved upward. Similarly, the pupils in the top group must have moved downward. That is, there is a negative relationship between initial standing and the direction in which change is most likely.

As an example of this effect, the pupils who stand highest on an achievement measure at the beginning of the school year will probably show little if any increase in score at the end of the year, and may even show a

decline. On the other hand, pupils who score lowest at the beginning of the year will probably show considerable increase. Educators have sometimes been misled by this effect and have assumed that their programs were more functional for low achieving pupils than for high achieving pupils, when in reality all that was involved was the regression effect (the statistical tendency for scores to move toward the average). Similarly, a group of pupils placed in a remedial program because they stand low on a pretest can be expected to show considerable improvement; but again the improvement may be spurious, as a consequence of the regression effect.

This problem creates real difficulties if pupils are tracked on the basis of fall scores and teachers are evaluated on the basis of change in achievement of their pupils. For example, assume that pupils are tested in reading in the fall and the lowest third are put in Miss Jones' class, the middle third in Miss Smith's class, and the highest third in Mrs. Williams' class. We can anticipate that at the end of the year Miss Jones' class will show much improvement and Miss Smith's will show modest gain, but Mrs. Williams will be fortunate if her pupils show any growth at all. The problem is that the gain the pupils show is materially affected by regression effect, so to evaluate the teacher on the basis of pupil gain would be manifestly unfair.

There are statistical procedures for attempting to eliminate this effect, but as Bereiter (1963) commented, it is impossible to be certain that appropriate adjustments have been made; and the expertise to do even the best that can be done with the problem is not widespread. And, of course, all the out-of-school influences on achievement standing discussed earlier also influence gain, although to a lesser degree. So it is clearly inappropriate to use pupil change as a way of evaluating teachers where a teacher may suffer as a consequence of the error involved.

Teacher Performance Tests. A procedure for evaluating teachers which attempts to bypass the problems of change is the performance test or the evaluative teaching unit (Flanders, 1974). In it, the teacher teaches a prescribed brief unit (sometimes as little as a few minutes or as much as two weeks) and pupil knowledge is then tested. The attempt is made to minimize the problems of measuring gain by teaching material in which pupils should have little or no preknowledge, so that all presumably start at "ground zero." But the other problems of using pupil achievement to evaluate teachers still apply. In addition, there are questions of whether teaching material which does not have to be integrated into previous knowledge requires the same skills as the usual teaching setting and whether such short-term learning generalizes to long-term learning. There is the final difficulty that the performance of teachers on a unit of a few minutes does not predict their performance on a two-week unit (McDonald, 1974). Assuming that either can be used to predict year-long performance then seems risky.

Even if the measurement of standing or gain in achievement were a satisfactory way of evaluating teachers, there is still the problem of selecting the objectives to be measured.

What Objectives Should Be Measured?

Although subject matter achievement has been the primary focus of the discussion thus far, it is clear that schools are charged with and have accepted some degree of responsibility for many other kinds of pupil growth.

The Need for Multiple Measures. Over a long period schools have given attention to the social development and the moral values of pupils. And a broad view of the relationship between school and society suggests that when a problem emerges in the society, one of the first steps is likely to be to

involve the school in solving the problem. Traffic problems led to driver education; a concern for the loyalty of government employees led first to a ban on teaching about communism in the schools and later to the requirement that it be taught; problems of drug abuse have led to drug abuse education in the schools; concern about sexual attitudes has led to sex education; concern for occupational choice has led to career education in the schools; and when concern for segregation of the races became pressing for the society, the first and the major attempt to deal with the problem was delegated to the schools. To evaluate teachers and schools solely on the basis of the subject matter gains made by pupils grossly under-represents the broad range of objectives for which teachers and schools have been given some degree of responsibility. Yet for many of these objectives there are no measures which are immediately, for some even remotely, available.

Simple Versus Complex Learning. Even within the subject matter realm there are problems which are largely ignored. One of these problems is the need to distinguish complex achievement growth from simple growth and to provide appropriate measurement for each. Memory of facts (rote memory) falls at the simplest level and complex problem solving, abstracting, and generalizing fall at the complex level. The distinction is between retrieving information (memory) and processing information in its varying degrees of complexity. There is some evidence from a number of studies that the teaching behaviors which are associated with greatest growth in simple tasks are different from those which are associated with greatest growth in complex tasks (Solomon, Bezdek, and Rosenberg, 1963; Soar, 1968; Soar and Soar, 1972, 1973).

Most studies of pupil achievement fail to make this distinction; and the current stress on criterion-referenced measurement, emphasizing "small-

step" learning, seems likely to focus on simple kinds of learning. Measures of complex learning are slow and difficult to construct, in contrast to measures of simple learning, which can be more easily and quickly developed. Evaluating all subject matter at all grade levels would almost certainly require the construction of many new measures which would likely emphasize simple kinds of achievement, given the ease with which they can be constructed and the emphasis on criterion-referenced measurement. If teachers were to be evaluated on the basis of pupil achievement, then, it seems likely that the teacher who emphasizes simple learning would be more positively evaluated than the teacher who emphasizes more complex learning. This would be an unfortunate result.

A further problem related to the difficulty of measuring complex achievement growth is the likelihood that some highly valued objectives grow too slowly to show change within a school year -- objectives such as complex problem-solving skills, citizenship, attitudes, learning to get along well with others, and creative expression. On the other hand, it seems likely that measures of short-term learning would tend to emphasize simpler kinds of learning.

Other Problems in the Use of Pupil Outcomes

A description of an application of accountability in England a century ago makes one of the problems clear (Small, 1972). In that setting, teachers were evaluated on the number of their pupils who attained the minimum level of achievement expected for the particular grade. The result was that teachers concentrated their efforts at the minimum level of proficiency, with a consequent lowering of the quality of instruction.

Another problem of serious consequence in the use of pupil measures

is raised by the OEO study of performance contracting, which found that the superior achievement of performance contracting programs disappeared when the teaching was controlled to eliminate the possibility of teaching the test (Page, 1972, 1973). The implication seems clear that, in a setting in which financial return follows from pupil achievement, teaching the test is likely to occur at least a portion of the time. This is a very reasonable finding and one which is well known, even in cases where a financial return is not involved -- teaching to the Regents Examination, for example.

A final problem is the possibility of bias if the teacher is the test administrator. Even outside test administrators have difficulty not "helping" pupils; but where a teacher is affected personally, it seems possible that his or her behavior might be influenced, even though unconsciously. This problem could be dealt with by using only specially trained test administrators, but this could be very costly.

Summary

When all these problems in the use of pupil achievement for teacher evaluation are considered, they become overwhelming. The influence of the teacher is minor compared to out-of-the-classroom influences -- pupil ability, previous knowledge, the home, the peer group, motivation, and others. What the pupil brings to the classroom in this respect is clearly a much stronger determinant of where he or she will stand at the end of the year than anything that has been done in the classroom. Influences on the development of future achievement measures seem likely to limit them to relatively simple measures for some time to come. Tests available for measuring the other objectives for which the teacher is to some degree responsible are relatively few. In addition to these problems, there are statistical difficulties in the measurement of

change which are extremely serious, if not disabling. They are still further exacerbated by the likely problems of teaching the test, of the teacher giving attention primarily to a small portion of the students, and of obtaining valid measurement in the classroom.

Taken all in all, this is an imposing array of difficulties, most of which have gone unrecognized when it is proposed that teachers be evaluated by measuring the outcomes of their pupils.

REFERENCES

Anderson, G. J. "Effects of Classroom Social Climate on Individual Learning." American Educational Research Journal 7: 135-53; March 1970.

Bereiter, C. "Some Persisting Dilemmas in the Measurement of Change." Problems in Measuring Change. (Edited by C. W. Harris.) Madison: University of Wisconsin Press, 1963.

Brophy, J. E. Stability in Teacher Effectiveness. R & D Report Series 77. Austin: Research and Development Center for Teacher Education, University of Texas, July 1972.

Cronbach, L. J. Essentials of Psychological Testing. Second edition. New York: Harper and Brothers, 1960.

Flanders, N. A. "The Changing Base of Performance-Based Teaching." Phi Delta Kappan 55: 312-15; 55: January 1974.

Garber, M., and Ware, W. B. "The Home Environment as a Predictor of School Achievement." Theory Into Practice 11: 190-95; June 1972.

Lord, F. M. "Elementary Models for Measuring Change." Problems in Measuring Change. (Edited by C. W. Harris.) Madison: University of Wisconsin Press, 1963. Chapter 2, pp. 21-38.

McDonald, F. J. "The State of the Art in Performance Assessment of Teaching Competence." Performance Education: Assessment. (Edited by T. E. Andrews.) Albany: Multi-State Consortium on Performance-Based Teacher Education, New York State Education Department, 1974.

Mayeske, G. W., and others, A Study of Our Nation's Schools. U. S. Department of Health, Education, and Welfare, Office of Education, Report No. DHEW-OE-72-142. Washington, D.C.: Government Printing Office, 1972.

Medley, D. M. "Research and Assessment in PBTE." AACTE Leadership Training Conference on Performance-Based Teacher Education. St. Louis, April 30, 1974.

Mosteller, F., and Moynihan, D. P. On Equality of Educational Opportunity. New York: Random House, 1972.

Page, E. B. "A Final Footnote on PC and OEO." Phi Delta Kappan 54: 575; April 1973.

_____. "How We All Failed at Performance Contracting." Phi Delta Kappan, 54: 115-17; October 1972.

Rosenshine, Barak. "The Stability of Teacher Effects Upon Student Achievement." Review of Educational Research 40: 647-62; December 1970.

Small, Alan A. "Accountability in Victorian England." Phi Delta Kappan 53: 438-39; March 1972.

Soar, R. S. and Soar, R. M. Classroom Behavior, Pupil Characteristics, and Pupil Growth for the School Year and for the Summer. Grant numbers 5 ROI MH 15891 and 5 ROI MH 15626, National Institute of Mental Health, U. S. Department of Health, Education, and Welfare. Gainesville: University of Florida, 1973.

_____. "An Empirical Analysis of Selected Follow Through Programs: An Example of a Process Approach to Evaluation." Early Childhood Education. (Edited by I.J. Gordon.) Seventy-first Yearbook, Part II, National Society for the Study of Education. Chicago: University of Chicago Press, 1972. Chapter 11, pp. 229-59.

Soar, R. S. "Optimum Teacher-Pupil Interaction for Pupil Growth." Educational Leadership Research Supplement 2: 275-80; December, 1968.

Solomon, D.; Bezdek, W. E.; and Rosenberg, L. Teaching Styles and Learning. Chicago: Center for the Study of Liberal Education of Adults, 1963.

Why Should All Those Students Take All Those Tests?

Published by the

NATIONAL EDUCATION ASSOCIATION
1201 16th Street, N. W., Washington, D. C. 20036

May 1975

WHY SHOULD ALL THOSE STUDENTS TAKE ALL THOSE TESTS?

(EVERY-STUDENT TESTING OR SAMPLING OF SELECTED GROUPS?)

The NEA Task Force on Testing, in its first interim report, states:

> The Task Force believes there is overkill in the use of standardized tests and that the intended purposes of testing can be accomplished through less use of standardized tests, through sampling techniques where tests are used, and through a variety of alternatives to tests....
>
> Representatives of the testing industry and others told the Task Force that sampling of student populations could be as effective as the blanket application of tests that is now so common. Some suggested that such procedures, in addition to increasing the assurance of privacy rights, would conserve time, effort, and financial expenditure.[1]

The blanket use of tests (every-pupil testing) in some state assessment and local testing programs appears to require inordinate amounts of time and resources on the part of teachers, other personnel involved in test administration and interpretation, and the students themselves.

Criticisms of the blanket use of tests have come from a variety of prominent researchers, evaluators, and other educators.

House, Rivers, and Stufflebeam, in their evaluation of the Michigan accountability system, concurred that in that state:

> Statewide testing as presently executed also raises the question of the feasibility of every pupil testing. This practice appears to be of dubious value when the cost of such an undertaking is compared with the resulting benefits

[1] In Task Force and Other Reports presented to the Fifty-Second Representative Assembly of the National Education Association, July 3-6, 1973, Portland, Oregon. pp. 26-46.

> to local level personnel.... The local, and hence overall, costs could be reduced by a matrix sampling plan which requires that each student tested take only a few items.... In the long run, a matrix sampling plan will be the only one feasible from a cost and time standpoint. The cost and time required for every pupil testing for the whole state would be horrendous.... We feel that it /strict adherence to a statewide testing model/ will result in useless expenditures of monies and manpower, in addition to producing unwarranted disruptions of the educational programs within a great number of schools.[2]

In a paper entitled "Criteria for Evaluating State Education Accountability Systems," the National Education Association has laid down fifteen basic principles, one of which is as follows:

> If the state desires test data for its own planning purposes, it should use proven matrix sampling techniques which will not reveal schools and which will greatly reduce costs.
>
> Matrix sampling techniques can give an accurate picture of the state by various categories much more efficiently than testing each child with an entire instrument.[3]

It was with such admonitions as these in mind that this paper was developed. And while some procedures are appropriate for evaluating all students in one way or another for particular purposes, it would appear that there is gross over-use of blanket testing procedures.

To help teachers and other educators better understand some main considerations related to sampling, the NEA obtained permission from Dr. Frank Womer, Michigan School Testing Service, University of Michigan, to reproduce

---

[2]House, Ernest R.; Rivers, Wendell; and Stufflebeam, Dan. An Assessment of the Michigan Accountability System. Michigan Education Association and National Education Association, March 1974. pp. 14-16.

[3]National Education Association. "Criteria for Evaluating State Education Accountability Systems." Washington, D. C.: the Association, n.d...

material from a monograph of his on developing assessment programs.[4] In addition, Dr. Womer prepared, especially for this paper, a section on item sampling. Dr. Womer's recommendations follow.

* *

Determining Whether Sampling Is To Be Used

The decision whether to test an entire population or use a sample involves a combination of concerns. Clearly there are policy considerations; clearly there are psychometric[5] considerations; clearly there are data collection considerations; and clearly there are cost considerations. The best possible staff and consultant thinking on this question should be brought to an advisory committee for them to consider very carefully.

Probably the most crucial consideration is a policy one, since psychometrics, data collection, and cost generally would argue on the side of sampling rather than using an entire population. If it is deemed wise for policy reasons to test all students in a population, that preference, typically, will have to be weighed against available resources and technology; so we will consider first the policy implications of the two choices.

One needs to look carefully at the purposes and goals of a specific assessment program in determining whether sampling is appropriate. If all of the specific purposes and objectives of an assessment program can be met by group results, then sampling must be considered.

---

[4]Womer, Frank B. Developing a Large-Scale Assessment Program. Denver: Cooperative Accountability Project, 1973.

[5]Editor's note: Psychometrics in the strictest sense of the definition has to do with the measurement of mental abilities. It has come to be used much more broadly to define a wide range of activities in assessment and evaluation.

The only assessment situation that clearly calls for common data collection on all members of the population is when it is deemed essential, for improved decision making, to have exactly the same test information for every pupil in a given grade in a state (or other assessment unit). It is exactly this situation that has prevailed for years in local school districts that have every-pupil achievement or ability testing at some grade level. Historically, the compulsory state testing programs were examples of this situation; the voluntary programs were not. If a state mandates common testing for all students it is taking over a role that local districts traditionally have held. This may be good or this may be bad depending on one's point of view of the role of a state department of education. It certainly has important policy implications.

There are many facets to this point, but it should be kept clearly in mind that it is not necessary to test every pupil at a given grade level on identical material in order to get a good picture of education outcomes of groups of students; it is necessary only if one feels that each teacher in an entire state at a given grade level must have the same information for each pupil.

Probably the greatest advantage of sampling is that for a given amount of effort (and money) one can gather more usable information than by using an entire population. If the goals of an assessment program are to gather statewide information only, it is hard to conceive of any reason for testing all students in a given grade. For example, if there are 50,000 third-graders in the state of Limbo, and one wants to gather state statistics only, it is very possible that a sample 5,000 students (or even 500) would

be sufficient if they are selected by a probability sample....[6] Or, if one can afford to test all 50,000 third-graders, and if it is deemed wise to do so, one could select ten 5,000-pupil samples and secure information on ten subject areas, or one could go into great depth of information gathering in two or three subject areas. The combinations of possibilities of sampling pupils and content are almost endless.

If one wants district-level information, then sampling becomes a different situation. In a school district with one third grade, sampling of pupils is hardly possible for most assessment purposes. In school districts with many third-graders, sampling could provide a greater variety of information than common testing on every pupil, in the same fashion as at the state level. Specific decisions of how far to carry sampling should be made only after advice from a sampling statistician. Sampling is a highly developed technical field, and the implications of any decisions to sample or not to sample must be reviewed by competent samplers.

Other "compromise" possibilities exist. One could test all students in a population with one short test, while using a sampling approach for other tests. This approach would provide some common information on all students but would allow for greater depth of data collection over a subject area.

> Principle: Sampling of pupils and/or content should be given very serious consideration for all large-scale assessment projects. The only situation where it may not be useful is one where it is deemed essential to collect common information on all students in a statewide population

---

[6]Editor's note: For information on probability samples, see Womer, op. cit.

of students. Sampling should be used to maximize the collection of usable information for stated assessment purposes at the lowest possible cost and effort.

* * *

Sampling with total tests is less complicated to administer, but since it is likely to be subject to error in administration and consequently less reliable, in some cases item sampling may be more useful. Therefore, Dr. Womer was asked to prepare an additional statement on the purposes and potential of item sampling. His statement follows.

Item Sampling

The process of item sampling in testing is more useful for one of two purposes:

1. to increase the amount of group test results that can be obtained from students in a given period of time; or
2. to decrease the amount of testing time necessary to obtain large amounts of group test information from students.

For either purpose, it is essential to keep in mind that item sampling is useful for gathering information about groups of students. Thus it is

a technique for use with relatively large groups, not a classroom-sized group or even three or four classes within a building.

Example 1

A school system has 500 students in the sixth grade. A standardized reading test is to be administered for a one-shot systemwide survey. The test takes 45 minutes to administer, which is all the time that can be taken from a busy schedule at the end of the year.

Staff are unhappy that only reading is to be surveyed. Some major changes were made in the mathematics curriculum three years before and they feel it would be valuable to survey this subject also. By randomly selecting only 250 of the students to take the reading test, the other 250 could be given a 45-minute mathematics test at the same time.

Example 2

A school system has 1,000 fourth-graders. It is desired to do an in-depth study of student outcomes for 100 different behavioral objectives in mathematics. Each objective requires the use of eight questions. The total of 800 questions would require one student to spend perhaps 15 hours of testing time to attempt all of them.

By randomly dividing up the objectives and items into five different subtests (each with 20 objectives and 160 items), each subtest could be administered to 200 students (randomly

selected). This would require only 3 hours of testing time per student (manageable) rather than 15 hours (unmanageable), and group results would still be available for all 100 objectives (800 items).

In either example the results will be usable for group analyses. Any slight reduction in accuracy due to sampling error is apt to be much less than errors due to increasing testing time of students beyond some reasonable amount. Systematic errors due to fatigue, disinterest, poor motivation, teacher concern, and other conditions of testing can easily outweigh a small sampling error.

The OEA Accountability Position Paper

An Overview

Accountability is an important but primitive concept as applied to education. Whereas some national leaders have accorded much importance to this area, it is by no means well developed and understood in education. Although definitions of accountability are plentiful, they vary widely. None has attained anything like widespread acknowledgement, let alone acceptance.

A sound educational accountability system must assess educational performance in relation to performance factors that are not only considered to be important by the constituency and participants in the educational system, but which can be demonstrated as such. Complicating the adoption of accountability is that these performance factors are inextricably tied to the performance of many groups who participate and are responsible for the total educational enterprise. It is thus vital to identify a set of accountability variables which encompass society's goals for education, society's investment in education, the operating characteristics of the educational system, and the nature of the societal context in which the educational system is embedded.

Whatever is done in the name of accountability in education shall be consistent with the frame of reference for teaching within which the teacher is operating. To demand that an educator be accountable for goals or outcomes that have not been mutually developed and accurately defined is not only grossly unfair to the educator but is destructive to humanistic goals and purposes.

Educators can and should be held responsible for being able to demonstrate the rational and professional bases for whatever they do, be it through research, logical thought, experience, consistency with theory or etc. This aspect of professional decisions is one for which all educators can and should be held accountable.

In our pluralistic society, the accountability system should promote diversity, not conformity. Opportunities for diversity must exist for the child, the teacher, the school and the community. Each entity has a right to be itself within the goal structure of Education and for an American democracy. In short, the accountability system should be responsive to individual differences that exist among children, parents, teachers, schools, and even communities. Ideally one would hope that such a system could even be personalized.

The formation of any local accountability system should be a joint effort negotiated among the board of education, administrative staff, and teachers' association, and should be subject to periodic review and updating.

LIMITING FACTORS WHICH INFLUENCE ACCOUNTABILITY IN EDUCATION

1. The school is rooted in the social and economic context of the community, state, and nation in which it is located. The school cannot escape that context. What the school can accomplish is simultaneously made possible and limited by the social and economic conditions which surround it.

2. The school is an expression of the values, competence, and wealth of the people who create it. It is shaped by those factors.

3. The same factors which shape the school -- that is, the values, competence, and wealth of the population -- also shape the non-school educative processes and conditioners of learning which take place in the home, neighborhood, community, state, and nation.

4. The school does not educate alone. The home educates; the neighborhood educates; the community educates; the state educates; and the nation educates.

5. The combined influence of the home, neighborhood, community, state, and nation is greater than the educational power of the school alone.

6. The school is not responsible and cannot be held accountable for all that the student learns. There are some kinds of learning for which the school is primarily responsible, some for which it shares responsibility with other institutions, and others for which it cannot be given responsibility and cannot be held accountable.

7. Many of the educative processes available to classroom teachers have been predetermined before classroom instruction begins. Teachers already have limits placed around what they can do by virtue of the time, space, equipment and materials, instructional methods, and curriculum content established by higher authorites as the framework for instruction. Teaching techniques are the joint product of decisions made by the federal government, the state legislature, the State Department of Education, the local district board, local school administrators, classroom teachers, and, to a degree, students and their parents. Teachers, therefore, are not alone responsible for teaching techniques and cannot be held solely accountable for what they accomplish.

8. Any system for accountability in education should acknowledge the many educative processes and conditioners of learning which are at work outside as well as inside the school.

9. The professional staff member may be held responsible for achieving reasonable objectives only when all other objectives upon which they depend have been met.

OEA - POLICY ON EDUCATIONAL ACCOUNTABILITY

The Ohio Education Association recognizes the current demand for accountability in education. The term "accountability" as applied in education is subject to varied interpretations and, therefore, the Association expresses the following beliefs as a policy statement.

I. State Level Responsibilities

The Association believes that accountability should improve and equalize educational opportunity; however, the present state of the art of educational evaluation is such that an effective method of accountability at the state or national level has not yet been developed.

The Association believes that present "national testing programs" and "statewide testing programs" are not valid for the comparison of students, buildings or districts. The process leads to narrowed instruction which may encourage educators to gear the curriculum to the testing program and to rating systems which are destructive to the educational process.

The Association believes that an extensive program of accountability will be of enormous expense which could divert needed resources from the instructional program. Should accountability monies be appropriated by the Legislature, the State Department of Education shall insure that the benefits derived from the educational accountability program more than offset the costs.

The Association believes that accountability will require collective bargaining procedures, statewide grievance guidelines and a professional practices board for successful implementation.

II. Local Level Responsibilities

The Association recognizes that schools are not the only "educators" in American society. Any accountability system for education must recognize local control of public education, must be implemented at the school district level and be based on the goals and objectives of that district. Unreasonable or unrealistic goals established can only lead to frustration and disappointment.

The Association believes there should be a shared accountability in education. The legislature, the school community, the parents, the board of education and the school administration must all be held accountable for providing: the necessary resources including adequate building, equipment and supplies; supplementary services; a reasonable teacher-pupil ratio; an acceptable educational atmosphere, including a community interest in education and parental cooperation in achieving educational objectives.

The Association believes that each educator should continue to be held accountable only for those outcomes over which he has reasonable control.

I. EVALUATION OF EDUCATORS

The most important purposes of evaluation are to improve the effectiveness of the individual practitioner, to inspire professional growth, and to shape a successful career in education. Therefore, evaluation should not be restricted to classroom teachers but must include professional personnel at all levels.

I. State Level Responsibilities

A. The state government has the responsibility to insure that local programs of educator evaluation provide due process to all educators and comprehensive and equitable evaluations.

B. All programs of evaluation of educators shall be locally developed and administered.

II. Local Level Responsibilities

A. The local board of education shall adopt negotiated programs and procedures for the evaluation of educators. These shall be mutually developed by, and be acceptable to, the teachers' association, the administration and the board of education.

B. The local teachers' or education association should use the following guidelines in developing the locally negotiated procedures for staff evaluations:

1. The program shall require that evaluations be carried on with the full knowledge and awareness of the educator.

2. The evaluator must meet with the educator to be evaluated before any evaluation procedures are initiated for the purpose of mutually determining the nature of, and the criteria for the evaluation and for providing background on those points to be evaluated.

   a. The evaluator, through the board of education, must provide the framework which offers encouragement and resources to enable the staff person to develop and implement any agreed upon improvements.

   b. The local evaluation program should recognize the need for a variety and quantity of evaluation techniques such as self-evaluation, classroom observation, job targets, and video taping.

   c. The program should recognize that a professional evaluative process emphasizes performance rather than personality traits.

d. Educators should recognize that the assessment of their performance may include input from any staff members whose working relationship qualifies them to make evaluation judgement.

e. The program shall prohibit the use of pupil testing as the major means of evaluating and/or ranking educators, pupils, groups of pupils by buildings, and school systems. When the individual practitioner being evaluated deems testing to be a desirable part of the evaluative process, the testing program shall meet scientific testing procedures including -- pre-test, post-test, control groups and valid, reliable comparison data.

3. The evaluator should meet promptly with the staff member being evaluated following any formal evaluation procedure. The meeting should be held for the purpose of discussing those points previously agreed upon.

4. The evaluator must give to the staff member a copy of the formal evaluation report and a sufficient amount of time before any conference is held, so that the staff person can study the evaluation thoroughly.

5. Every educator shall have the right to indicate those evaluative documents in his personal file which he believes are obsolete or otherwise inappropriate to retain. After a joint review with the superintendent or his designee, materials deemed obsolete should be destroyed. Disputes over the retention of such documents should be considered grievances, with action beginning at the superintendent's level.

6. The educator shall have the right to institute a grievance if the staff member concludes that deviation from agreed-upon procedure has occurred.

7. The procedure should provide that those educators whose performance is constantly judged to be unsatisfactory may, after reasonable time and resources for improvement have been provided, be subject to dismissal after following the due process procedure.

8. The programs and procedures for evaluation of educators should be periodically reviewed and, if necessary, revised by a joint association, administration and board of education committee.

9. The professional staff member may be held responsible for achieving reasonable objectives only when all other objectives upon which they depend have been met.

II. EVALUATION OF STUDENTS

I. State Level Responsibilities

A. The appropriate state agency should protect the student from the invasion of privacy due to the release of achievement information beyond the professional staff directly responsible for the education program.

B. The appropriate state agency should protect against the compiling and release of group data which when released could adversely affect the district, building, class, or any other identifiable group.

II. Local Level Responsibilities

A. The local board of education has the responsibility to protect the student from the invasion of privacy due to the release of achievement information.

B. The local board of education should establish a policy that will protect the staff from releasing any unauthorized evaluation data.

C. There should be student evaluation at the local level that is based on a comprehensive approach which would include, but not be limited to:

1. Reports of student progress such as grades, checklists, pass-fail options, parent conferences, etc.

2. Interviews between -- pupil-teacher, parent-teacher, teacher-other support professionals, and teacher-student.

3. The testing at the local level should be for such functions as placement, pupil inventory, and achievement assessment.

4. Student observations and self-evaluations should use a variety of activities, such as oral and written evaluations, and audio and video taping.

III. TESTING PROCEDURES

I. State Level Responsibilities

A. The state shall provide a public educational system that will insure broad opportunities for all school age children, and, therefore, it shall not establish nor permit the establishment of testing programs which force students, teachers, and local schools into rigid patterns of development.

B. If statewide evaluation data becomes necessary, the appropriate state agency has the responsibility to insure that:

1. Only proven matrix sampling techniques shall be used,

2. Neither individual nor school data will be revealed, and

3. Costs will be kept at reasonable levels.

II. Local Level Responsibilities

A. The local board of education shall adopt a policy that provides for an effective local program of testing and this testing program must be negotiated by and approved by the teaching staff.

B. The local association should use the following guidelines in developing the local testing procedures that will be negotiated:

1. The development of a testing program that is in response to local school district needs.

2. The testing program should prohibit the release of specific individual test scores, such as: a grade placement, percentile, or percentile rank.

3. The testing program shall prohibit the release of group test results to the general public and prohibit the comparison of buildings or classes.

4. If any test results are released to a student's parents they should be accompanied by a full description of the test, its instructional purposes, and its limitations.

5. All aspects of the testing program should be continually monitored by a committee composed of a cross section of the teaching staff. This would require an annual evaluation and readoption of the testing program by the local school staff.

IV. ADMINISTRATIVE ACCOUNTABILITY

I. State Level Responsibilities

The appropriate state agency shall provide for (1) a minimum standard of administrative certification, such standards to be regularly reviewed; and, (2) the funding and broad goals of a preparation program for school administrators.

II. Local Level Responsibilities

A. The local board of education shall adopt job descriptions for every administrative position which have been jointly agreed upon by the board of education and the professional staff.

B. Each job description shall include a clear statement of goals and objectives.

1. Such objectives should be reasonable and definitive in nature and in terms of outcomes for each administrative level within the local school structure.

2. There must be prior agreement by all parties in regards to the objectives and their means of evaluation.

3. The professional staff member may be held responsible for achieving reasonable objectives only when all other objectives upon which they depend have been met.

V. STAFF INVOLVEMENT IN CURRICULUM DECISIONS

I. State Level Responsibilities

A. The appropriate state agency has the responsibility of establishing state minimum curriculum requirements.

B. In developing and establishing state minimum curriculum requirements, the appropriate state agency shall involve all professional groups of educators within the state.

C. The state minimum curriculum requirements shall reflect only those objectives common to the electorate of the State of Ohio.

D. The appropriate state agencies shall require and provide financial support to school districts to develop and implement broad curricula beyond the state's minimum requirements.

II. Local Level Responsibilities

A. The local board of education shall have the responsibility of adopting a negotiated policy on the curriculum after opportunities for specific staff-community involvement have been provided including at least the following guidelines:

1. The curriculum established within each school should be reflective of both the district and individual building's written philosophy and written general instructional objectives.

2. Such philosophies and general instructional objectives should have been developed with community involvement in an effort to reflect local needs.

3. The development of specific departmental curricula shall be the responsibility of those professional staff members working in that department or grade level.

4. The local Board of Education shall provide within its policies a statement on controversial issues in the curriculum. Such a policy should insure academic freedom and provide for safeguards for the professional staff and students.

5. The components for an effective framework for developing the curriculum should include the following:

a. a curriculum council -- consisting of teachers, principals, and supervisors representative of the entire staff and acting as the clearing house for the various programs to effect curriculum improvement.

b. curriculum development coordinator(s) -- whose job it is to coordinate and give leadership and assist the curriculum committees in reaching their goals.

c. curriculum committees -- provided to carry out various agreed study projects.

d. staff resources -- other necessary components for curriculum development must include staff resources provided by the local school board such as: opportunities for staff participation in workshops; seminars; conferences; school-visitations; and, direct resource assistance from school personnel or outside consultant services.

e. curriculum review -- there shall be maximum opportunities for continuous review of the school curriculum.

VI. INSERVICE AND PROFESSIONAL DEVELOPMENT

I. State Level Responsibilities

A. The appropriate state agency shall provide continuous financial support for the development and implementation of an effective inservice program at the local level.

B. Inservice funds must be adequate to provide for sufficient staff development and implementation of mutually desired programs at the local level.

II. Local Level Responsibilities

A. Every local board of education shall adopt a negotiated policy to provide for a regular inservice program which develops ongoing professional growth opportunities for teachers, administrators, and school board members. Such a program should incorporate the following provisions:

1. It should be flexible to reflect needs of the community as well as the individual needs of the educators within that school district.

2. It should permit the initiation of the program by any party of the educational community such as the teacher's association, the local administration, or a State Professional Practices Board.

3. It should be structured with a provision for a reward mechanism to promote participation and involvement.

4. It should specify the person or persons who will be primarily responsible for the institution and development of the inservice program.

B. The local school district shall make commitments in areas of budget allocations and staff time to effect quality inservice programs. Such commitments shall include the following provisions:

1. Educators shall be provided with released time for inservice programs.

2. Local pay schedules shall include recognition of professional growth for all related experiences.

3. Local boards of education shall pay the cost of professional study related to the implementation of new programs.

VII. FINANCIAL IMPLICATIONS OF ACCOUNTABILITY

I. State Level Responsibilities

A. The state department of education should not spend monies for educational accountability that have not been specifically designated and appropriated by the state legislature for that purpose.

B. Should educational accountability monies be appropriated by the state legislature, the state department of education shall insure that the benefits derived from such an accountability program more than offset the costs for such an activity.

II. Local Level Responsibilities

The exact dollars to be spent on educational accountability locally shall be negotiated regularly by the local board of education and the local teacher's association.

CONCLUSION

To satisfy the present demands for accountability, to help the schools change, and to be of least damage to what the schools are doing well, local association leaders should follow the guidelines of this document. It should be recognized that the guidelines provide some minimal considerations on each particular issue and that additional guidelines may need to be developed at the local level. The statements and seven sections in this accountability document are all tied together and must not be separated. The activities specified must be viewed as operational objectives for all local associations and will only be implemented through negotiations at the local level.

The local accountability process must provide for a thorough and continuous informational program that permits the community to be aware of the progress as well as the needs of the district. Communications addressed to the public should be prepared in simple, readily understood language. Educational jargon and vagueness should be avoided. The importance of working with the mass media is critical in communication with the public and candor in all communications generally leads to improved relationships. It is further recommended that schools make copies of their personnel and curriculum policies available to the general public by placing them on file with public libraries within the district.

If accountability measures are to be adopted, teachers must work to have them take the form of local negotiations, which are more adaptable than legislation and easier to change and modify as new circumstances arise. Teachers must strive to be recognized and accepted as the source of authority and decision making in curriculum and instruction.

Finally, teachers--through their associations--can take appropriate collective action in response to the implementation of ill-advised accountability measures.